IOWA STATE UNIVERSITY

ZFNGenome: A gBrowse-based tool for identifying Zinc Finger Nuclease target sites in model organisms

ZINC FINGER CONSORTIUM

Deepak Reyon[1], Jessica R Kirkpatrick[1], Jeffry D Sander[2], J Keith Joung[2,3], Daniel F Voytas[1,4], Clark R Coffman[1] & Drena Dobbs[1]

[1] Dept. of Genetics, Development & Cell Biology, Bioinformatics & Computational Biology Program, Iowa State University; [2] Center for Cancer Research, Center for Computational & Integrative Biology, Massachusetts General Hospital; [3] Dept. of Pathology, Biological & Biomedical Sciences Program, Harvard Medical School; [4] Dept. of Genetics, Cell Biology & Development, Center for Genome Engineering, University of Minnesota

ABSTRACT

Zinc Finger Nucleases (ZFNs) have tremendous potential as tools to facilitate genomic modifications, such as precise gene knockouts or replacement, and to advance both basic research and clinical applications, including gene therapy. ZFNGenome is a comprehensive resource for locating ZFN target sites in the genomes of model organisms.
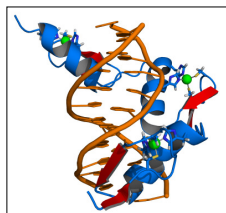
C2H2 zinc fingers form a 3-finger array that binds DNA

• **ZFNGenome** is a gBrowse-based tool for identifying and visualizing target sites for OPEN-generated ZFNs in model organisms, including yeast, *Arabidopsis*, soybean, *Drosophila*, and *C. elegans*. Target sites in several other model organisms, including human, mouse, and rat, will be included in the near future.

• **ZFNGenome** provides researchers with information about each target site, such as its chromosomal location, position relative to transcription initiation site(s), and expected *in vivo* activity, with a confidence score generated by ZiFOpT, a Naïve Bayes classifier for predicting *in vivo* functionality of ZFN target sites.

• Users can query **ZFNGenome** using several different criteria, e.g., gene ID, transcript ID, or target site sequence.

• Based on **ZFNGenome** analysis, OPEN-generated ZFPs are expected to successfully target 25,174 out of 29,291 (86%) of protein coding transcripts in zebrafish, 41,944 out of 45,767 (92%) transcripts in human and 30,193 out of 39,640 (76%) in *Arabidopsis*.

1 – Oligomerized Pool Engineering (OPEN)

• **OPEN** is a protocol developed in the Joung Lab for efficient construction of Zinc Finger Proteins with high affinity and specificity for a given 9-mer target site (Maeder et al., 2008. *Mol Cell* 31:294).

• OPEN-generated ZFPs have been used to perform successful genomic modifications in several organisms, including: zebrafish, tobacco, human, and *Arabidopsis* (Foley JE, et al., 2009. *PLoS ONE* 4:e4348; Townsend JA, et al., 2009. *Nature* 459:442. Feng Zhang, et al., PNAS 2010).

Figure from: Maeder ML, et al., 2008

2 – ZiFOpT

• **ZiFOpT** is a Naive Bayes classifier capable of distinguishing *active* vs *inactive* ZFP target sites (Sander & Reyon, submitted).

• An **active ZFP target site** is a 9-mer DNA sequence for which at least one ZFP has been obtained (using the OPEN protocol) that produces >3-fold activation of β-galactosidase activity in a bacterial 2-hybrid (B2H) *in vivo* functional assay.

• ZiFOpT was trained and tested on a dataset of 135 experimentally validated ZFP target sites.

• We evaluated performance of ZiFOpT using a standard leave-one-out cross-validation strategy. Results (below) indicate that ZiFOpT can reliably distinguish active *vs* inactive target sites with ~87% accuracy.

Classifier	Correlation Coefficient	Accuracy %	Specificity[+] %	Sensitivity[+] %	ROC AUC
Naïve Bayes	0.61	87	0.91	0.92	0.87

3 – ZFNGenome Features

ZFNGenome
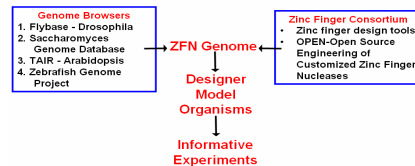
Goal: Create a user friendly interface between sequenced and annotated genomes of model organisms and the resources of the Zinc Finger Consortium.

• **Features:**
  • All potential OPEN target sites are identified and displayed.
  • Genomic features including Exons, Introns and Untranslated regions are displayed.
  • Links to *ZiFiT* and *ZiFDB*:
    • *ZiFiT* http://bindr.gdcb.iastate.edu/ – A server that aids design of ZFPs by identifying potential target sites in any given DNA sequence and providing links to reagents for building the corresponding ZFPs.
    • *ZiFDB* http://bindr.gdcb.iastate.edu:8080/ZiFDB/ – A comprehensive repository for experimental data for all publicly available ZFPs and ZFNs.
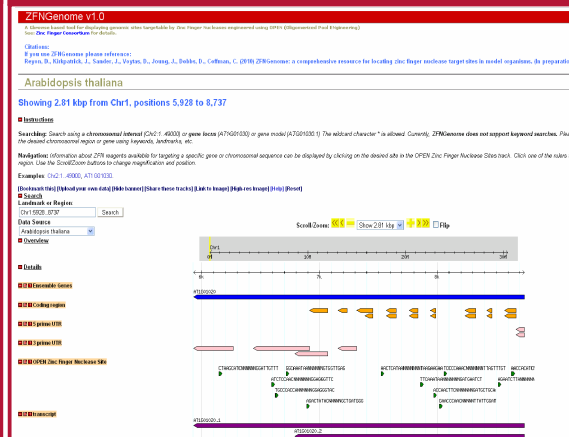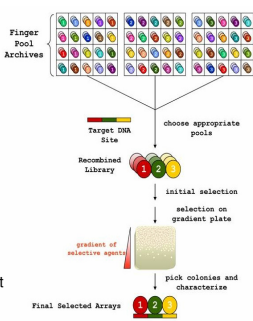 • Identify off-target sites using organism-specific BLAST searches.
  • Export ZFN target sites to other gBrowse based browsers such as TAIR and Flybase.
• **Future directions:**
  • Include additional model organisms.
  • Identify and incorporate sites that can be targeted using other zinc finger engineering methods (modular assembly).

4 – ZFNGenome Output

• We scan the entire genome of a model organism and find all 9-mer DNA sequences that are targetable using currently available OPEN reagents.

• All potential target sites are identified and stored in a MySQL database, along with other relevant information, including: start position, Gene ID, Transcript ID, feature (Intron, Exon, Coding Exon, etc.).

• To facilitate examination of **ZFNGenome** output, the chromosomal context of all identified ZFN target sites can be analyzed using an integrated gBrowse based genome browser.

SUMMARY

ZFNGenome

• is designed to help researchers identify the "best" target sites for sequence-specific Zinc Finger Nucleases generated using OPEN design procedures.

• employs ZiFOpT, a machine learning algorithm, to identify the target sites most likely to function *in vivo*.

• allows users to analyze target sites in their chromosomal context via a gBrowse based genome browser.

• currently provides a complete compendium of ZFP target sites in yeast, *Arabidopsis*, soybean, *Drosophila,* and *C. elegans*. We intend to incorporate rice, zebrafish and human in the near future.

• All information is freely available and accessible at: *http://bindr.gdcb.iastate.edu:8888/ZFNGenome*